

REMARKS

This application contains claims 1-3, 5-7, 9-18, 21-24, 26-31 and 33-35. Claim 9 has been canceled without prejudice. Claims 1, 10, 15, 22 and 29 are hereby amended. No new matter has been added. Reconsideration is respectfully requested.

Claim 15 was objected to for an informality. The claim has been amended to correct the informality as suggested by the Examiner.

Claims 1-3, 9-13, 22-24, 26, 29-31 and 33 were rejected under 35 U.S.C. 103(a) over Halstead, Jr., et al. (U.S. Patent 5,963,893) in view of Oflazer et al. ("Morphological Disambiguation by Voting Constraints"). Applicant has amended independent claims 1, 22 and 29 in order to clarify the distinction of the present invention over the cited art. These claims recite a method, software product and apparatus for morphological disambiguation of an input string, based on generating candidate analyses of the string and then selecting one or more of the analyses based on the relative frequency of occurrence of the linguistic patterns.

The claims have been amended to state that the linguistic patterns are evaluated using a statistical base, which is created by morphologically analyzing a corpus of text. The statistical base is built by finding the relative frequencies of occurrence of the linguistic patterns of the words in the corpus, independently of their lemmas. This added limitation is based on original claim 9 (now canceled) and on the method shown in Fig. 3 and described on page 16, lines 4-29, of the present patent application.

Halstead describes a word breaking facility for identifying words within a Japanese text string based on morphological processing (abstract). As explained in response to the previous Official Action in this case (and acknowledged by the Examiner in the present Official Action), Halstead does not teach or suggest determining relative frequencies of occurrence of linguistic patterns independent of the lemmas to which the patterns are applied.

Oflazer describes a morphological disambiguation system in which different morphological rules (or "constraints") are used to parse a sentence. After all the

applicable rules have been applied to a given sentence, the constraints “vote” in order to choose the parse that best matches the sentence (page 222, col. 2, lines 3-6). Oflazer’s parsing method is thus essentially context-dependent. The rules are manually programmed, as are the voting weights that are assigned to the rules (page 225, col. 1, lines 8-20). Oflazer indicates that it is “conceivable that votes can be assigned or learned by using statistics from disambiguated corpora,” but this sort of statistical voting is left “for future work” (page 224, col. 1, last paragraph).

In other words, although Oflazer suggests that a statistical base could be used in determining voting weights, he gives no instructions as to how such a statistical base might practically be built. Rather, he indicates that future work would be required to implement such a method. The only practical direction Oflazer gives in this regard is that the statistics should be gathered from disambiguated corpora, i.e., from documents in which the pattern and lemma of each word have been resolved.

Oflazer thus teaches away from the limitation, stated in amended claims 1, 22 and 29, that the statistical base is built by finding the relative frequencies of occurrence of the linguistic patterns of the words in the corpus, independently of their lemmas. The benefit of building the statistical base in this manner is that no prior disambiguation of the corpus is required, i.e., the statistics can be gathered from any corpus of text without prior processing of the corpus. This feature of the present invention is surprising in that it permits an individual word to be disambiguated using statistics derived from a corpus in which the meanings of the words remain ambiguous.

Amended claims 1, 22 and 29 are therefore believed to be patentable over the cited art. In view of the patentability of these independent claims, dependent claims 2, 3, 10-13, 23, 24, 26, 30, 31 and 33 are also believed to be patentable.

Claims 5-7 and 14 were rejected under 35 U.S.C. 103(a) over Halstead in view of Oflazer and further in view of Zamora (U.S. Patent 4,862,408). These claims depend from claim 1. In view of the patentability of claim 1 as amended, claims 5-7 and 14 are believed to be patentable, as well.

Claims 15-18, 21, 27, 28, 34 and 35 were rejected under 35 U.S.C. 103(a) over Zamora in view of Halstead and further in view of Oflazer. Applicant respectfully traverses this rejection.

Claims 15, 27 and 34 recite a method, apparatus and software product for searching a corpus of text, wherein candidate analyses of words in the corpus are selected based on the relative frequency of occurrence of their respective patterns independent of the lemmas to which the patterns are applied. The lemmas of the selected analyses are entered in an index of the corpus, to which a search query may then be applied.

Zamora describes a method for analyzing text using a paradigm. He creates a file structure in which each word in a list of words (or “dictionary”) is associated with a set of paradigm references. These references generate all forms of each of the lemmas of the words in the list (abstract). In other words, Zamora uses all possible linguistic forms of each of the words in a given list (col. 2, lines 66-68), without discriminating between the more and less frequent forms, as required by claims 15, 27 and 34.

As pointed out above and acknowledged by the Examiner, Halstead does not teach or suggest determining relative frequencies of occurrence of linguistic patterns independent of the lemmas to which the patterns are applied.

Furthermore, as explained in reference to claims 1, 22 and 29, Oflazer teaches away from finding or using relative frequencies of occurrence of the linguistic patterns of the words in a corpus independently of their lemmas. Oflazer suggests only that statistics may be assembled from disambiguated corpora (page 224, col. 1, last paragraph).

Thus, claims 15, 27 and 34 are believed to be patentable over the cited art. In view of the patentability of these independent claims, dependent claims 16-18, 21, 28 and 35 are believed to be patentable, as well.

Applicant has studied the additional reference made of record by the Examiner (Ezeiza et al., “Combining Stochastic and Rules-Based Methods for Disambiguation in Agglutinative Languages”). Although Ezeiza makes mention of

Serial No. 09/606,326

Customer No. 07278

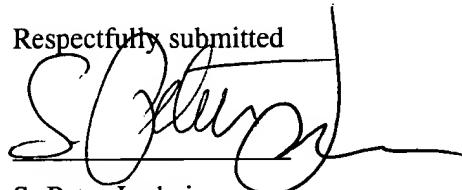
Attorney Docket No. 06727/000H370-US0

stochastic methods, he neither teaches nor suggests the idea of finding or using relative frequencies of occurrence of the linguistic patterns of the words in a corpus independently of their lemmas. Therefore, the claims in this application are believed to be patentable over Ezeiza, as well.

Applicant believes the amendments and remarks presented hereinabove to be fully responsive to all of the objections and grounds of rejection raised by the Examiner. In view of these amendments and remarks, Applicant respectfully submits that all of the claims in the present application are in order for allowance. Notice to this effect is hereby requested.

Date: November 28, 2005

Respectfully submitted



S. Peter Ludwig

Reg. No. 25,351

Attorney for Applicants

DARBY & DARBY, P.C.  
P.O. Box 5257  
New York, NY 10150-5257  
212.527.7700